

EFFECTIVENESS OF ST. JOHN'S WORT (*HYPERICUM PERFORATUM*) VS. PLACEBO  
OR OTHER ANTIDEPRESSANTS IN THE TREATMENT OF MILD TO MODERATE  
DEPRESSION: A META-ANALYSIS OF RANDOMIZED CONTROL TRIALS

By

James Benjamin Hooker  
MPH Candidate  
UNC School of Public Health

Master's Paper

Dr. Russ Harris

Dr. David Collier

## ABSTRACT

**Background:** Depressive disorders, including mild to moderate depression, are common illnesses that affect an estimated 1 in 5 individuals and account for substantial societal and individual costs. Cost is one barrier to an individual seeking treatment for depression with standard antidepressants. A more affordable antidepressant could lessen this barrier to the treatment of mild to moderate depression. Inexpensive extracts of the flowering plant *Hypericum perforatum* (St. John's wort or SJW) have long been used as an alternative treatment for depression. While the popularity of SJW has been growing in the US, there is still doubt as to its effectiveness. The purpose of this review is to examine the effectiveness of SJW alone compared to placebo or to other antidepressants in the treatment of mild to moderate depression.

**Methods:** I conducted a literature search for randomized clinical trials of human subjects published in English between 1998 and 2004 using MEDLINE, BIOSIS v3.0, Alt Health Watch, AMED, and CINAHL. I used the key words "antidepress\*," "depress\*," "hyperic\*," "john's wort," "johns wort," "johnswort," "blind\*," "random\*," "clinical trial," "rct," and "placebo." To be included in the analysis, a study must: 1) randomize subjects, 2) compare SJW alone to either placebo or other antidepressant, 3) enroll subjects  $\geq 18$  years old, 4) use the Hamilton Depression Rating Scale (HAM-D) to measure outcome, 5) study patients with mild to moderate depression, and 6) define "response" as one whose HAM-D score decreases by  $>50\%$  from baseline or is  $\leq 10$  at the end of the trial. I excluded studies with subjects under 18 years old and those that used a different definition for "response," did not compare SJW to placebo or other antidepressant, or included subjects with depressive disorders other than mild to moderate depression. Once I extracted relevant data from each trial, I performed a meta-analysis using the MIX 1.7 program.

**Results:** Of 68 articles identified through the literature search, 12 studies met *a priori* criteria for inclusion in the analysis. Five studies compared SJW to placebo, 5 studies compared SJW to other antidepressants, and two studies compared SJW to both placebo and another antidepressant. Five of 7 studies reported SJW was significantly more effective than placebo in treating mild to moderate depression. These 7 studies included 1510 subjects, lasted 6 to 8 weeks, and used doses of SJW ranging from 500 to 1500-mg per day. Tests for heterogeneity among studies were not significant. The test for publication bias was also not significant. Meta-analysis produced pooled outcomes measure for three comparisons. Significant difference in response was found between SJW and placebo, but not between SJW and other antidepressants, or SJW and SSRI.

**Discussion:** Results of this analysis of 12 studies suggests that treatment with SJW is significantly more effective than placebo and equivalent to other antidepressants, and SSRI's, in treatment of adults with mild to moderate depression. The strength of these conclusions is lessened by the small sample of included studies. As the popularity of SJW continues to grow, further studies will be necessary to determine a standard dosing regimen and to study adverse effects and to better study effectiveness. There is not currently enough of a database of adequate trials to draw final conclusions on SJW effectiveness versus placebo or other antidepressants. In spite of these limitations, SJW potentially offers a less expensive alternative to standard antidepressant medications.

## INTRODUCTION

### *Depression*

Unipolar depressive disorders include major depression, which can be classified as mild to severe, and dysthymic disorder.<sup>1</sup> In any given year, depressive disorders affect about 9.5% of the general population 18 and older and it is estimated that one-third of individuals will suffer at least one depressive episode during their lifetime. As of 1990, unipolar major depression accounted for 10.7% of years lived with disability worldwide, 6.1% of all disability adjusted life years (DALY) in developed nations, and 10.3% of DALY in persons aged 15 to 44 worldwide. By 2020 unipolar depression is projected to be second only to ischemic heart disease as a cause of disability worldwide. Dysthymic disorder, a chronic and mild form of depression, is more prevalent than unipolar depression with a lifetime prevalence of 5.4% among US adults 18 and older. An estimated 10.9 million persons are affected by dysthymia each year, with an estimated 9.9 million affected by unipolar depression. About 2 in 5 of those diagnosed with dysthymia will be meet criteria for major unipolar depression later in life.<sup>2</sup> Any estimate of prevalence of depressive disorders or of their economic impact is likely an underestimate since depression often goes undiagnosed and unreported.<sup>3</sup> Even those who go undiagnosed, however, will likely seek treatment for their symptoms of depression.

The majority of those with a diagnosable depressive disorder have dysthymic disorder, a form of depression that is mild and chronic. These patients are more likely to use complementary and alternative medicines (CAM) in an effort to manage their condition without doctor's visits or prescription medications. Although CAM therapies already enjoy widespread use, it is important to gather evidence to support, or refute, their effectiveness. It is at least as

important to collect and analyze data on adverse effects from CAM therapies as it is to establish whether or not they are effective.

### *Complementary Treatments for Depression*

The Cochrane Collaboration defines complementary and alternative medicine (CAM) as “diagnosis, treatment, and/or prevention which complements mainstream medicine by contributing to a common whole, by satisfying a demand not met by orthodoxy, or by diversifying the conceptual frameworks of medicine.”<sup>4</sup> The popularity of CAM therapies has grown recently, with as many as 40% of adults using at least one CAM therapy for at least one year as of 1998.<sup>5</sup> The popularity of CAM therapies can be attributed to the patient’s desire to find treatments with fewer adverse effects as well as to the perception that alternative therapies are less authoritarian and more empowering, allowing the patient to take more control over their treatment.<sup>6</sup>

Depression ranks as one of the more popular conditions for which patients choose a CAM therapy and a variety of therapies have been cited as useful in treatment of depression. These treatments include therapies focusing on movement, such as applied kinesiology and exercise, and those that employ touch, like acupuncture and massage. Hypnotherapy uses hypnosis to help relieve depressive symptoms, while Naturopathy focuses on enhancing the body’s own ability to heal itself. Other philosophies hold that disease states can be manipulated by simple elements like oxygen or color or that illness can be treated through principles of faith. Remedies based on herbs or plants include aromatherapy, medical herbalism, and phytomedicine.<sup>6</sup> Relaxation,

exercise, and herbal treatments, such as St. John's wort, rank as the most often utilized CAM therapies in the treatment of depression.<sup>4</sup>

### *History of St. John's Wort*

*Hypericum perforatum*, a perennial herb found growing in Europe, Western Asia, North Africa, and North America, is a shrub whose yellow blossoms produce a sap when crushed. This sap is also known as Andro Haimon, or Mar's blood, due to its blood-like appearance and consistency.<sup>7</sup> *Hypericum perforatum* gained part of its popular name from its tendency to flower in late June, around the birthday of St. John the Baptist, as well as from the tradition of gathering the herb's yellow flowers as part of the feast to celebrate St. John.<sup>8,9</sup> The moniker of St. John was combined with "wort," the Old English word meaning plant, to give the full popular name of St. John's wort.<sup>8,10</sup>

History is ripe with examples of *Hypericum*'s use in treating conditions such as depression, anxiety, and insomnia.<sup>8,11,12</sup> The medical school of Salerno lists "herba demonis fuga" (herb that chases away the devil) in its 13<sup>th</sup> century list of medicinal plants. In 1525, Paracelsus reports using *Hypericum* for the treatment of depression, melancholy, and over-excitability (most likely anxiety). Early 17<sup>th</sup> century Franciscan monks used "fuga demonum" (the devil's scourge) to ward off evil spirits and demonic possession.<sup>8,10</sup> The above examples demonstrate a long-standing belief that *Hypericum* has properties effective in treating mental illness.

St. John's wort has enjoyed its most widespread use as an antidepressant medication in Germany gaining both the official approval of the German government and the largest share of

the antidepressant market.<sup>12</sup> In Germany, sales of St. John's wort exceed sales of all other antidepressants combined with 20 times more prescriptions written for St. John's wort than for Prozac.<sup>13,14</sup> In spite of its popularity and well-established use in parts of Europe, questions remain concerning both the effectiveness and safety of the herb.

### *Previous Studies*

Past clinical trials have attempted to demonstrate the herb's effectiveness in treating depression. Serious methodological flaws weakened conclusions of earlier trials primarily conducted in Germany. Failure to use intention-to-treat analysis or to use statistical tests of significance to compare responder rates of the treatment group to rates of the placebo group are two of the more common flaws of these early studies.<sup>8</sup> Many studies were too short to allow for meaningful comparison.<sup>12</sup> Studies intending to compare St. John's wort to other antidepressants have often failed to use a high-enough dose of the antidepressant to allow for meaningful comparisons with the herbal treatment.<sup>12</sup> Failure to use DSM-IV criteria to properly identify mildly to moderately depressed subjects may constitute the most serious methodological flaw of earlier studies. Many of these studies used subjects whose diagnoses were more consistent with acute stress disorder or adjustment disorder with depressed mood. Symptoms caused by these conditions were more likely to fluctuate or spontaneously resolve over the course of a trial when compared to symptoms caused by a primary depressive disorder.<sup>12</sup> Results of trials conducted with such serious methodological flaws are of little value in evaluating St. John's wort's effectiveness or safety.

Popularity of St. John's wort has been growing in the U.S. primarily through promotion by popular media outlets. In addition to the media's portrayal of St. John's wort as safe and effective, characteristics of the U.S. population have also helped drive the herb's popularity. Many believe that herbal products are more natural and therefore safer than other pharmaceuticals. Some find the fact that no prescription is required for St. John's wort an attractive quality. By avoiding an interaction with a mental health professional and the pharmacist, these individuals can avoid any personal acknowledgement of mental illness, thus avoiding the societal stigma of mental illness.<sup>12</sup> Regardless of the reason, use of St. John's wort by Americans has been on the rise, demonstrated by recent increases in sales of the supplement as well as by the availability of St. John's wort in retail outlets and through internet vendors.<sup>15, 16</sup>

### *Placebo Controversy*

Over time, studies have shown that there is a substantial and widely variable placebo effect operating among patients being treated for depression. Recent research has also indicated that this placebo effect may be increasing in magnitude.<sup>17</sup> With such a strong placebo effect confounding studies of novel antidepressants, it becomes necessary to compare any new treatment to both placebo and an active comparator, such as an SSRI or TCA. For example, if a *Hypericum perforatum* extract is compared only to an SSRI, then the study may prematurely conclude that the effects of *Hypericum* are equivalent to the SSRI. On the other hand, if *Hypericum* is compared only to placebo, then the study is more likely to conclude that *Hypericum* is not more effective than placebo.<sup>18</sup>

The existence of such a powerful placebo effect demands a certain progression of studies before one can draw conclusions concerning the effectiveness of a novel therapy. Written in 2002, the National Manic-Depressive Association Consensus Statement on the use of placebo in clinical trials of mood disorders states that a new drug must first be found to be more effective than placebo before studies between the new drug and standard treatments can conclude the novel treatment is effective.<sup>19</sup> In order to accurately assess *Hypericum*'s effectiveness, one must first determine whether or not it is significantly more effective than placebo. Should these studies show *Hypericum* is more effective than placebo, focus can shift to whether or not *Hypericum*'s effects are equivalent to those of standard antidepressants. In addition, it is important that this second group of studies also include a placebo group to properly account for placebo effect.<sup>19</sup>

Two studies to date have compared a *Hypericum perforatum* extract to both placebo and an active comparator. Philipps and associates compared the effects of *Hypericum* to both placebo and Imipramine in the treatment of moderately depressed subjects.<sup>20</sup> This group used the HAM-D to measure change in symptoms over an 8-week period in patients randomly assigned to treatment with a maximum dose of 100-mg Imipramine, *Hypericum* extract containing 0.2 to 0.3% hypericin, or placebo. *Hypericum* was shown to be more effective than placebo and at least as effective as the dose of Imipramine used in the study. Given its positive effects and favorable side effect profile, this group concluded that *Hypericum* was effective, safe, and improved symptoms of moderately depressed subjects.<sup>20</sup> The Hypericum Depression Trial Study Group (HDTSG) compared the effects of an extract of *Hypericum* containing 0.12 to



0.28% hypericin to Sertraline and placebo. This group did not find any differences between *Hypericum* and placebo or between *Hypericum* and Sertraline.<sup>21</sup>

### *Public Health Issue*

While the FDA regulates labeling and restricts the types of claims that can be made about dietary supplements, manufacturers are currently not required to prove effectiveness or safety of products either before or after releasing them to market.<sup>22-27</sup> Since supplements do not undergo pre-market testing or post-market surveillance, it is important for independent groups to more closely study both St. John's wort's effectiveness and safety profile using methodologically sound techniques. It is the purpose of this paper to systematically review studies on the effectiveness of St. John's wort in treating mild to moderate depression. A meta-analysis of relevant studies will also be conducted, if appropriate.

## METHODS

### *Search Methods*

I identified relevant articles using a systematic search strategy limited to randomized control trials published in English language journals. Articles published from January 1998 to October 2004 were retrieved from online PubMed, BIOSIS v3.0, Alt Health Watch, AMD (Alternative Medicine), and CINAHL databases. I chose this date range based on trends noted in the published literature. Studies of SJW published prior to 1998 have been criticized for their methodological weakness. These methodological weaknesses include 1) used of inexperienced investigators, 2) failure to use a standardized instrument to measure depressive symptoms, 3) use of non-standard diagnostic criteria leading to heterogeneous study groups, and 4) short study duration.<sup>28</sup> In addition, studies prior to 1998 were conducted primarily in Europe and published in lesser-known journals. Studies published since 1998 have attempted to address the aforementioned methodological weaknesses and have more often been published in more respected journals.

Key words used in the initial search included “antidepress\*,” “depress\*,” “hyperic\*,” “john’s wort,” “johns wort,” “johnswort,” “blind\*,” “random\*,” “clinical trial,” “ret,” and “placebo.” Medical subject headings (MeSH) terms used in the search included “antidepressive agents,” “depressive disorders,” “depression,” and “Hypericum.” I have summarized search terms used in this search below in Table 1. I consulted a research librarian both during and following the search to confirm that I had conducted a complete and systematic search.

**Table 1.** List of terms, key words, and limits used in the systematic search of the literature for studies comparing SJW to either placebo or other antidepressants for the treatment of adults with mild to moderate depression.

[MeSH Terms]	[Key Words]	[Search Limits]
Antidepressive agents	antidepress*	blind*
Depressive disorder	depress*	random*
Depression	hyperic*	placebo*
Hypericum	john's wort	clinical trial
	johns wort	rct
	johnswort	

#### *A priori criteria for identifying studies*

I evaluated all potentially relevant articles identified through the literature search using *a priori* inclusion criteria. Only articles that met all criteria would be included in the analysis. To be included, studies must 1) be randomized, 2) compare SJW alone to either placebo, another antidepressant, or both, 3) enroll subjects older than 18 years of age, 4) utilize the Hamilton Depression rating scale (HAM-D) to measure clinical outcome, 5) study patients with mild to moderate depression, and 6) define “response” as a greater than or equal to 50% decline from baseline HAM-D score or a HAM-D score of less than 10 by the end of the study. Inclusion criteria are listed below in Table 2.

**Table 2.** *A priori* criteria used to identify articles for inclusion in comparison of effectiveness of SJW in treating mild to moderate depression versus placebo or other antidepressant.

- 1) Randomized allocation of subjects,
- 2) Compare SJW alone to placebo, other antidepressant, or both,
- 3) Enroll subjects greater than 18 years old,
- 4) Use HAM-D to measure clinical outcome,
- 5) Study patients with mild to moderate depression,
- 6) Define “response” as  $\geq 50\%$  decline from baseline HAM-D or final HAM-D  $< 10$ .

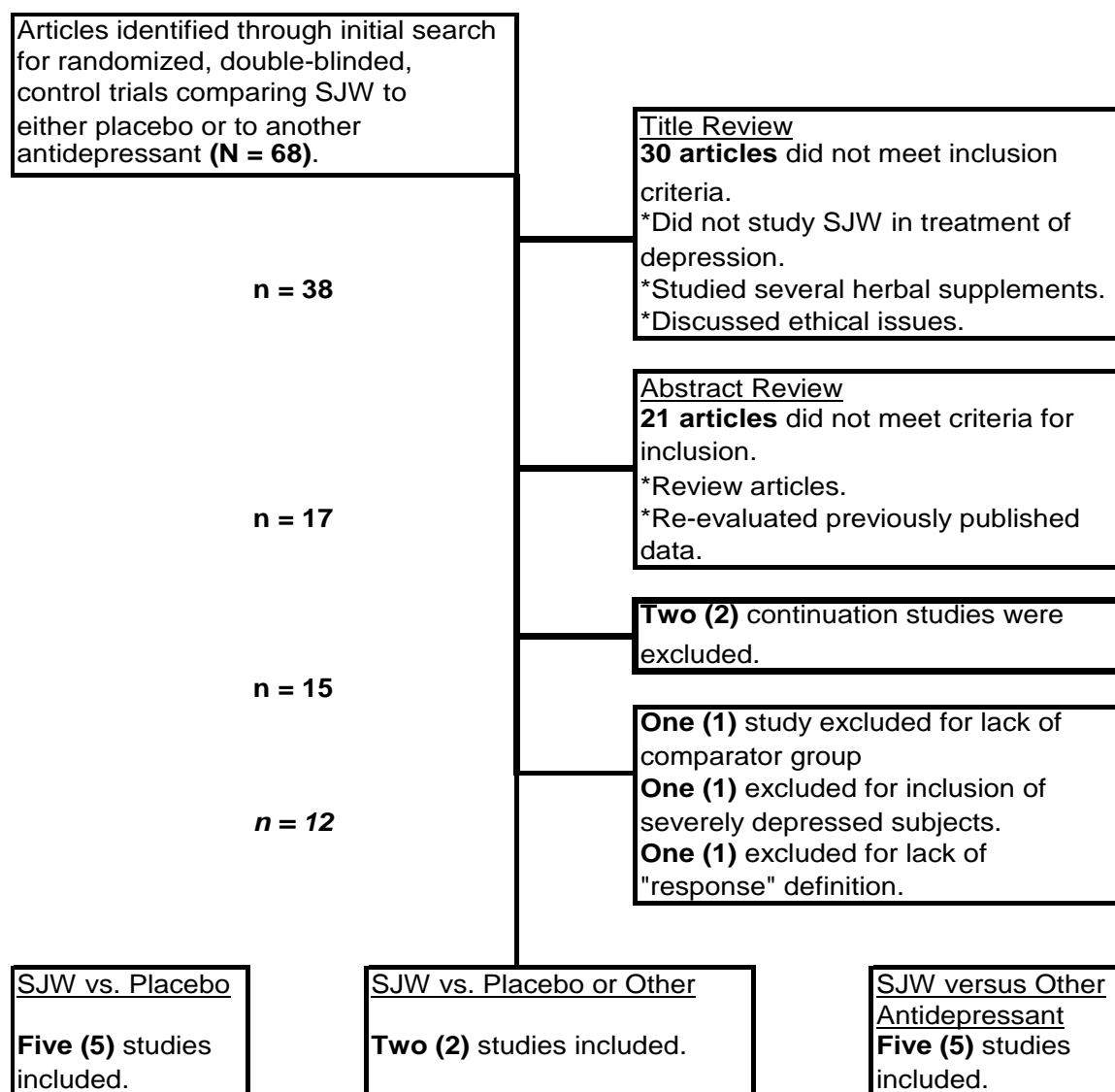
According to the DSM-IV, mild to moderate depression is defined by number of criteria symptoms present, severity of symptoms, and degree of dysfunction resulting from the condition. To be diagnosed with depression, a patient must exhibit a depressed mood and/or marked decrease in enjoyment from or interest in activities they normally enjoy. Additionally, the patient must exhibit three or more of the following: 1) weight loss or gain, 2) sleeping more or less than usual, 3) psychomotor retardation, 4) fatigue, 5) feelings of guilt or worthlessness, 6) decreased ability to concentrate, and 7) recurrent thoughts of death or suicide. Symptoms must be present for at least two weeks for the diagnosis to be made.<sup>29</sup>

Mild depression is diagnosed when five or six depressive symptoms (depressed mood and/or anhedonia with 3-4 other symptoms) of mild intensity are present for a two-week period. While mildly depressed subjects require substantial effort to navigate normal activities, these persons suffer only mild functional impairment. Moderate depression is defined by the presence of a few depressive symptoms in excess of the minimum needed for diagnosis. Moderate depression may also be diagnosed with minimum symptoms, but only if those symptoms are of moderate intensity and lead to greater functional impairment than seen in mild depression.<sup>29</sup> Mild to moderate depression is further described as uncomplicated by such co-morbidities as substance abuse, eating disorders, cognitive dysfunction, panic disorder, bipolar disorder, personality disorder, psychosis, or post-traumatic stress disorder.<sup>28</sup> Therefore, I excluded any studies that included patients with these complications from this analysis. I chose the definition for “response” to treatment based on its standard use in clinical trials of the effectiveness of antidepressants in treating depression.<sup>30</sup>

For use in research, DSM-IV definitions of mild to moderate depression are translated into a specific score on an instrument used to measure depressive symptoms. The 17-item HAM-D (HAM-D 17), the instrument used most often to measure primary end-points in antidepressant trials, has a maximum value of 52. “Severe” depression is defined by a score or equal to or greater than 28. A score from 16 to 27 defines “Moderate” depression and a score from 8 to 15 represents “mild” depression. “Remission” is defined as a score equal to or lesser than 7.<sup>30</sup>

### *Article Selection*

I reviewed randomized trials comparing the effectiveness of SJW to that of placebo or other antidepressants in treating adults (18 years or older) with mild to moderate depression. Only studies that used HAM-D scores to measure the primary outcome, “response” to treatment, were included. Following the literature search described above, I systematically reviewed titles and abstracts of articles and excluded articles that did not meet *a priori* criteria for inclusion. Had time and resources allowed, a second reader would have independently reviewed the articles identified through the literature search. Any disagreements over an article's inclusion would then have been adjudicated by evaluation of the complete article by both reviewers. After more extensive review, the reviewers would then come to a mutual decision on the article's inclusion or exclusion. Please see Figure 1 for a summary of article selection.



**Figure 1.** Flow diagram of literature search and application of *a priori* inclusion criteria to identify studies for systematic review.

### *Internal Validity*

Any study that failed to meet all of the stated inclusion criteria, or that reported data from an earlier study, was excluded from further analysis. I then reviewed studies that met inclusion criteria and judged the internal validity of the methods used by the following criteria: 1) was

subject randomization successful, 2) did randomized subjects meet the definition of “mild to moderate” depression according to baseline HAM-D score, 3) was intention-to-treat (ITT) method used in analysis, 4) did randomized subjects represent a typical primary care population, 5) were study subjects blinded, and 6) did loss to follow-up or crossovers significantly affect groups measured at the study’s endpoint. Criteria used to assess internal validity of included studies are listed below in Table 3.

**Table 3.** Criteria used to assess internal validity of articles included in comparison of effectiveness of SJW to that of placebo or other antidepressants in treating adults with mild to moderate depression.

- 1) Were study subjects effectively randomized?
- 2) Do randomized subjects meet definition of “mild to moderate” depression according to baseline HAM-D score?
- 3) Was intention-to-treat (ITT) analysis used?
- 4) Do randomized subjects represent a typical primary care population or are they more representative of a selective group?
- 5) Were subjects blinded to their group allocation?
- 6) Did loss to follow-up or crossover significantly affect groups when measured at the study’s endpoint?

Once assessed for internal validity, I assigned each study a grade of “good,” “fair,” or “poor.” If a study met all validity criteria, then it was deemed “good.” Studies meeting most criteria, and free of any major flaws, were graded as “fair.” Major flaws included 1) ineffective randomization, 2) inclusion of a highly specific group of subjects, 3) failure to use ITT analysis, 4) failure to blind subjects to group, and 5) patient attrition so great that it affects the final analysis. Studies found to have a major flaw were graded “poor” regardless of their adherence to other criteria. Had time and resources not limited this study, a second reader would have independently evaluated internal validity of included studies according to the criteria listed

above. Once the second reader had graded each article, the two readers would have met to discuss any differences in grading. Once identified, grading differences would then have been resolved through consensus.

### *Analysis*

Prior to statistical analysis, I performed a narrative analysis of the included papers. I started by critically reading each paper with a focus on the methods section. Close reading of each included study allowed me to perform a detailed comparison of studies that compared SJW to placebo and, then, SJW to other antidepressants. Data used for comparison can be seen in appendices 1 and 2. Had time and resources allowed, a second reader would have independently reviewed each paper and any discrepancies addressed.

In addition to systematically reviewing each study, I analyzed data from included studies to generate pooled outcome measures describing the effectiveness of SJW versus that of placebo and other antidepressants. I extracted experimental and reference data from each included study and placed this data into one of four categories. For studies comparing SJW to placebo, these categories included SJW responders, SJW non-responders, Placebo responders, and Placebo non-responders. For studies comparing SJW to other antidepressants, these categories included SJW responders, SJW non-responders, Other antidepressant responders, and Other antidepressant non-responders. I also made a third comparison that included SJW responders, SJW non-responders, SSRI responders, and SSRI non-responders. I compared SJW to Placebo to determine if SJW had any positive effect in the treatment of depression while accounting for the placebo effect. I compared SJW to Other antidepressants and SSRI to determine if treatment of depression with



SJW was equivalent to that of the other antidepressants. Had time and resources allowed, a second reader would have independently reviewed abstracted data and any discrepancies would have been addressed with the first reader.

I used an Excel-based meta-analysis program, MIX 1.7, to perform all calculations and defined statistical significance using an alpha level of 0.05.<sup>31, 32</sup> With the appropriate data extracted from included studies, I first evaluated included studies for statistically significant heterogeneity using the Cochran homogeneity statistic. If there is no significant heterogeneity found, then it is appropriate to perform a meta-analysis. Analysis of the Cochran result also allowed me to identify any important differences between studies so that these may be investigated further. To facilitate such analysis, the MIX 1.7 program produces a funnel plot comparing study effects versus different aspects of the included studies (sample size, p-value, or standard error). If appropriate, pooled outcome measures [risk ratios with 95% confidence intervals (CI)] will be reported for each comparison and presented in the form of a forest plot. The Mantel-Haenszel method will be used for these calculations and use of the fixed or random effects model will be based on whether or not there is statistically significant heterogeneity among included studies. I will then analyze forest plots for each comparison to draw conclusions on the effectiveness of SJW compared to placebo and other antidepressants in treating mild to moderate depression.

## RESULTS

Initial literature searches identified 68 citations. A review of titles of these 68 articles revealed that 30 did not meet inclusion criteria. Excluded articles were those that did not study the effects of SJW on mild to moderate depression, did not use human subjects, reviewed several herbal supplements used in treating depression (including SJW), or discussed ethical issues surrounding the use of placebo in clinical trials of antidepressants. Closer examination of the abstracts of the remaining 38 articles identified 17 articles that warranted closer consideration for inclusion. The 21 articles excluded during this examination consisted of review articles and papers that re-evaluated data published in earlier studies. Of the 17 articles identified for closer evaluation, two were excluded because they were continuation studies and not studies of treatment for acute disease. Another single article was excluded because, even though it studied three different concentrations of SJW, it had neither a placebo group nor an antidepressant group for comparison. Two other studies were excluded, one because it did not define “response” and the other because it included severely depressed subjects. The 12 remaining studies were included in the final analysis (please refer to Figure 1).

Of the 12 articles in the final analysis, 5 compared SJW to another antidepressant without using a placebo group. One study used Imipramine as the active comparator, one used Paroxetine, two used Sertraline, and four used Fluoxetine. Two studies compared SJW to both placebo and an active comparator. One such study compared SJW to placebo and Imipramine while the other compared SJW to placebo and Sertraline. Five studies compared SJW alone to placebo, without using an active comparator. For this analysis, data from the studies that compared SJW to both placebo and an active comparator was used in both comparisons, SJW

versus placebo and SJW versus other antidepressant. Data from seven studies was included in the comparison of SJW versus placebo and data from 7 studies was used in comparing SJW to other antidepressants.

### *SJW versus Placebo*

Seven studies were analyzed in making the comparison between SJW and placebo.<sup>9, 33, 20, 34, 28, 21, 35</sup> The studies by Shelton, et al<sup>28</sup> and the HDTSG<sup>21</sup> were conducted in the US while the others were conducted in Germany and France. Five of the 7 studies defined mild to moderate depression at baseline according to DSM-IV criteria.<sup>21, 28, 33, 34, 35</sup> Groups led by Schrader<sup>9</sup> and Philipp<sup>20</sup> used ICD-10 codes to define baseline depression. Schrader's group also used the 21-item HAM-D to measure outcome while the other studies used the 17-item HAM-D. Included studies randomized a total of 1510 subjects, lasted from 6 to 8 weeks, and tested 6 different preparations of SJW. Doses of SJW used in these trials varied from 500-mg to 1500-mg per day divided into one to three daily doses. Two of these studies also included another antidepressant arm.<sup>20, 21</sup> Philipp's group compared SJW to placebo and Imipramine and the HDTSG compared SJW to placebo and Sertraline. One study, led by Laakmann, included two active arms, each using a different preparation of SJW extract.<sup>33</sup> Data from this study was divided so that each extract could be compared to placebo individually. Details of the studies included for the comparison of SJW to placebo can be found in Table 4 and Appendix 1.

**Table 4.** Randomized, double blind trials comparing SJW to placebo, other antidepressant, or both.

Study	N	DSM IV	Baseline HAM-D			End	Hypericum extract	Dose (mg)	Other	Dose (mg)
			SJW	Placebo	Other		Preparation		Drug	
Schrader (1998)	162	No	20.13	18.76		6 w	Ze 117	500		
Laakmann (1998)	98	Yes	20.3	21.1		7 w	WS 5573	900		
Laakmann (1998)	98	Yes	20.9	21.1		7 w	WS 5572	900		
Philipp (1999)	263	No	22.7	22.7	22.2	8 w	STEI 300	1050	Imipramine	100
Kalb (2001)	72	Yes	20.1	19.7		6 w	WS 5572	900		
Shelton (2001)	200	Yes	22	23		8 w	LI 160	900- 1200		
HDTSG (2002)	340	Yes	23.1	22.7	22.5	8 w	LI 160	900- 1500	Sertraline	50- 100
Lecrubier (2002)	375	Yes	21.9	21.9		6 w	WS 5570	900		
Harrer (1999)	149	No	16.6		17.18	6 w	LoHyp-57	800	Fluoxetine	20
Brenner (2000)	30	Yes	21.3		21.7	7 w	LI 160	900	Sertraline	75
Schrader (2000)	240	No	19.65		19.5	6 w	Ze 117	500	Fluoxetine	20
Woelk (2000)	324	No	22.4		22.1	6 w	Ze 117	500	Imipramine	150
Behnke (2002)	70	No	20		20.7	6 w	Calmigen	300	Fluoxetine	40

**Notes:**

- All studies that did not use DSM-IV criteria to define depression used appropriate ICD-10 codes.
- Both studies involving Schrader used the 21-item HAM-D. All others used the 17-item HAM-D.
- Laakmann, et al (1998) included two active arms, each testing a different Hypericum extract. These arms are presented separately.

*SJW versus Other Antidepressant*

Seven studies were analyzed in making the comparison between SJW and other antidepressants.<sup>20, 21, 36, 37, 38, 39, 40</sup> Studies by the HDTSG<sup>21</sup> and Brenner, et al<sup>37</sup> were conducted in the US while the others were conducted in Germany. Five of the 7 studies defined mild to moderate depression at baseline according to ICD-10 codes while only Brenner, et al<sup>37</sup> and

HDTSG<sup>21</sup> used DSM-IV criteria. The study by Schrader<sup>38</sup> used the 21-item HAM-D to measure outcome while the other studies used the 17-item HAM-D. Included studies randomized a total of 1416 subjects, lasted from 6 to 8 weeks, and tested 5 different preparations of SJW. Doses of SJW used in these trials varied from 300-mg to 1500-mg per day divided into one to three daily doses. Groups led by Philipp and Woelk tested SJW versus the tricyclic antidepressant Imipramine.<sup>20, 39</sup> Five studies compared SJW to a selective serotonin reuptake inhibitor (SSRI), with two studies using Sertraline<sup>21, 37</sup> and three studies using Fluoxetine.<sup>36, 38, 40</sup> Imipramine doses used ranged from 100-mg daily to 150-mg daily. Sertraline doses used ranged from 50-mg to 100-mg per day and Fluoxetine doses used ranged from 20-mg per day to 40-mg per day. Details of the studies included for the comparison of SJW to other antidepressants can be found in Table 4 and Appendix 2.

#### *SJW versus Placebo and Other Antidepressant*

Two studies were identified that compared SJW to both placebo and another antidepressant.<sup>20, 21</sup> The HDTSG conducted its work in the US while the group led by Philipps performed their study in Germany. The HDTSG<sup>21</sup> used DSM-IV criteria to define mild to moderate depression at baseline while the Philipps group used ICD-10 codes.<sup>20</sup> Both studies used the 17-item HAM-D to measure outcome. These studies randomized a total of 603 subjects that were split into two comparisons (SJW versus placebo, SJW versus other antidepressant) and included in separate comparisons. Both studies lasted 8 weeks, but tested different preparations of SJW against different antidepressants. Doses of SJW used in these trials varied from 900-mg to 1500-mg per day divided into one to three daily doses. Philipp and associates tested SJW

versus placebo and versus 100-mg of Imipramine daily.<sup>20</sup> The HDTSG studied SJW versus placebo and versus 50 to 100-mg of Sertraline daily.<sup>21</sup> More detail on the studies included for the comparison of SJW to other antidepressants can be found in Appendices 1 and 2.

### *Other Characteristics*

Close evaluation of included studies revealed other differences between studies. Five of the included studies used a placebo run-in period for all groups prior to initiation of placebo or intervention.<sup>33, 34, 28, 21, 35</sup> In the studies that compared SJW to placebo, the placebo response rate ranged from 13.5% to as high as 63%. Three studies did not standardize the dose of SJW used according to amount of active ingredient.<sup>36-38</sup> The remainder of the included studies did take steps to standardize the dose of active ingredient used in the SJW arm. There also exists some between groups differences in the level of depression [low moderate (HAM-D 16 to 20) versus high moderate (HAM-D 20 to 24)] measured at baseline. Details of these other study characteristics can be seen in Table 5 below.

**Table 5.** Other characteristics of included studies.

Study	Placebo	Placebo	Standard	Baseline HAM-D	Baseline HAM-D
	run-in?	response	SJW dose	16-20 (low moderate)	20-24 (high moderate)
Schrader, et al (1998)	No	13.50%	Yes	placebo	Ze 117
Laakmann, et al (1998) a	Yes	32.70%	Yes		placebo WS 5573
Laakmann, et al (1998) b	Yes	32.70%	Yes		placebo WS 5572
Philipp, et al (1999)	No c	63%	Yes		placebo STEI 300 Imipramine
Kalb, et al (2001)	Yes	42.90%	Yes	WS 5572	placebo
Shelton, et al (2001)	Yes	18.60%	Yes		placebo LI 160
HDTSG (2002)	Yes	43.10%	Yes		placebo LI 160 Imipramine
Lecrubier, et al (2002)	Yes	42.30%	Yes		placebo WS 5570
Harrer, et al (1999)	No	None	No	Lo-Hyp-57 Fluoxetine	
Brenner, et al (2000)	No c	None	No		LI 160 Sertraline
Schrader (2000)	No c	None	No	Ze 117 Fluoxetine	
Woelk (2000)	No	None	Yes		Ze 117 Imipramine
Behnke, et al (2002)	No	None	Yes		Calmigen Fluoxetine

**Notes:**

a, b. These are the two active arms of Laakmann, et al (1998). Each arm tests a different Hypericum extract.

c. Philipp, et al (1999); Brenner, et al (2000); and Schrader (2000) used a "wash-out" period prior to initiating treatment or placebo.

*Internal Validity*

Studies were assigned a grade of good, fair, or poor based on their adherence to *a priori* criteria for internal validity (see Table 3 above). Of the studies that compared SJW to placebo,

four were judged to be good studies. Laakmann, et al <sup>33</sup> satisfied all criteria. Studies led by Shelton,<sup>28</sup> HDTSG,<sup>21</sup> and Lecrubier <sup>35</sup> included subjects that scored above 24 on the baseline HAM-D, and did not fit the definition of mild to moderate depression. In their study, the HDTSG set out to address methodological flaws identified in past studies.<sup>21</sup> This group took extra measures to assess their own methods including statistical analysis to measure the effectiveness of subject blinding. Their analysis concluded that blinding was significantly less effective in the Sertraline group, likely secondary to medication side effects. In addition, the HDTSG group had less than 80% follow-up, though statistical analysis revealed that dropout was evenly distributed across groups. None of the other studies in this group employed methods to assess effectiveness of blinding or to evaluate loss of subjects to follow-up. In spite of the aforementioned issues, the study by HDTSG was graded a good study.

Two studies in this group were deemed fair. Philipp and associates earned a fair grade by including subjects that did not meet the definition of mild to moderate depression at baseline and by not reporting how they blinded subjects to study group.<sup>20</sup> The study by Kalb, et al was judged as fair in spite of the fact that it met all criteria used to judge internal validity. This study's conclusions were weakened by a small sample size and short duration.<sup>34</sup> One study in this group, by Schrader and associates, was considered to be a poor study. This study suffered a fatal flaw when randomizing subjects to different study groups. The HAM-D scores of the SJW were found to be significantly higher than those of the placebo group when measured at baseline.<sup>9</sup>

Among studies that measured the effectiveness of SJW versus another antidepressant, three were judged as good. The study by HDTSG may have included subjects that did not meet the definition of mild to moderate depression and may not have adequately blinded subjects to



allocation, but the study was still strong. The HDTSG made efforts to evaluate the methods used to blind subjects and further investigate loss to follow-up were representative of a methodologically rigorous study.<sup>21</sup> The study by Woelk was good in spite of including subjects that did not meet the definition of mild to moderate depression and greater loss to follow-up in the Imipramine group.<sup>39</sup> Sample size and evaluation of the remainder of study method were considered rigorous enough to earn a good grade. Schrader met all criteria set forth to evaluate internal validity.<sup>38</sup>

Three studies that compared SJW to another antidepressant were judged to be fair. Philipp and associates included subjects that did not meet the definition of mild to moderate depression and did not report the method used to blind subjects to study group.<sup>20</sup> Brenner, et al did not use a standard pill size across study groups rendering blinding ineffective. This study also suffered from greater than 20% loss of subjects at follow-up.<sup>37</sup> Behnke, et al suffered from differential loss to follow-up, with the SJW group losing more than the Fluoxetine group, and failed to report the method used to blind subjects to group.<sup>40</sup> Harrer, et al was found to be a poor study. This study also failed to use a standard pill size across groups, thus negatively affecting blinding. In addition, the Fluoxetine group lost more than twice as many subjects to follow-up and the population studied was not representative of a primary care population.<sup>36</sup> Please see Table 6 below for a summary of Internal Validity assessment.

**Table 6.** Assessment of internal validity in included studies.

Study	Random	Mild-Mod Depression	ITT	Prim. Care	Subjects blinded	Sig. loss to f/up	Study Grade
Schrader (1998)	No a	Yes	Yes	Yes	Yes	No	poor
Laakmann (1998)	Yes	Yes	Yes	Yes	Yes	No	good
Philipp (1999)	Yes	No	Yes	Yes	?	No	fair
Kalb (2001)	Yes	Yes	Yes	Yes	Yes	No	fair
Shelton (2001)	Yes	No b	Yes	Yes	Yes	No	good
HDTSG (2002)	Yes	No b	Yes	Yes	No d, j	No k	good
Lecrubier (2002)	Yes	No b	Yes	Yes No	Yes	No	good
Harrer (1999)	Yes	Yes	Yes	c	No e	Yes f	poor
Brenner (2000)	Yes	Yes	Yes	Yes	No e	Yes g	fair
Schrader (2000)	Yes	Yes	Yes	Yes	Yes j	No	good
Woelk (2000)	Yes	No b	Yes	Yes	Yes j	Yes h	good
Behnke (2002)	Yes	Yes	Yes	Yes	?	Yes i	fair

**Definitions:**

"Mild to moderate" depression = baseline HAM-D 17-item score of 16 to 24.

Effects of loss-to-follow up = Distribution of subjects lost from each group.

Randomization = Assessed reported method and baseline group characteristics.

Blinding of subjects = Assessed by evaluation of reported method.

**Notes:**

a. Schrader (1998) SJW group had baseline HAM-D scores higher than placebo

b. Philipp (1999); Shelton (2001); HDTSG (2002); Lecrubier (2002); Woelk

(2000) included subjects who scored above 24 on baseline HAM-D

c. Harrer (1999) enrolled elderly subjects ages 60 to 80 years old

d. HDTSG (2002) analyzed effectiveness of blinding, found blinding was less effective in Sertraline group

e. Neither Harrer (1999) nor Brenner (2000) used standard pill size.

f. Harrer (1999) fluoxetine lost 23.5% to follow-up, SJW lost 11.6%

g. Brenner (2000) had less than 80% follow-up

h. Woelk (2000) imipramine lost 17.9% to follow-up, SJW lost 9.6%.

i. Behnke (2002) fluoxetine lost 8.6% to follow up, SJW lost 17.1%.

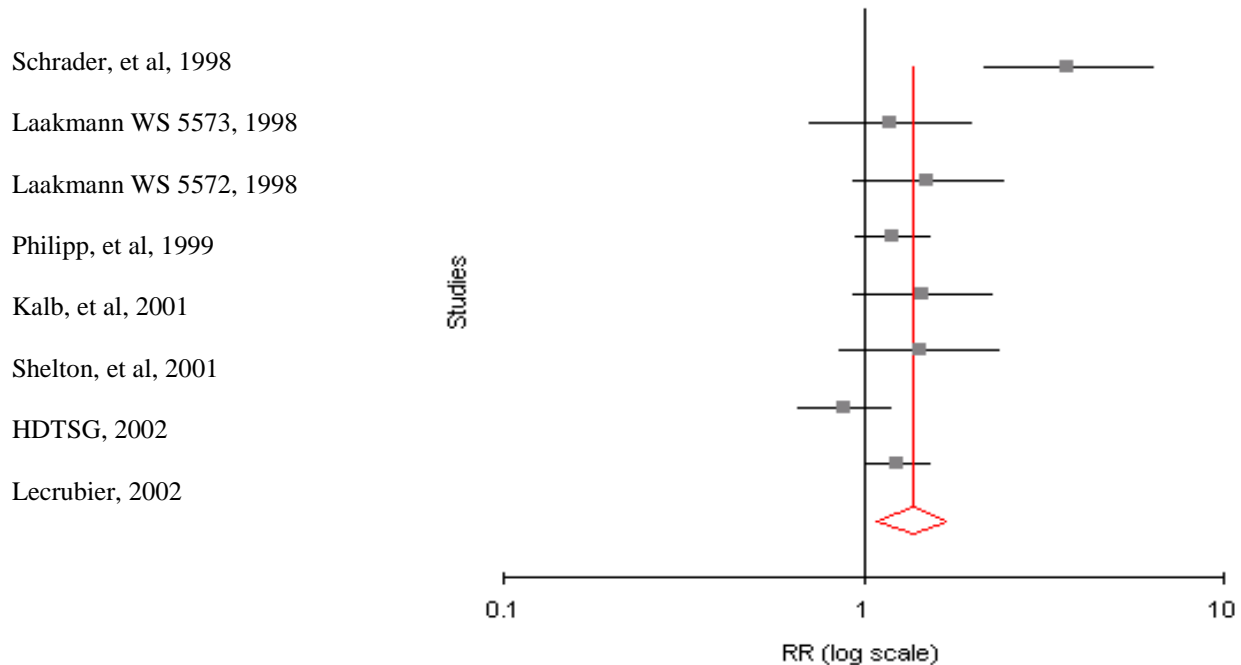
j. HDTSG (2002); Schrader (2000); and Woelk (2000) used double-dummy method to ensure subjects were blinded to treatment.

k. HDTSG (2002) had less than 80% follow-up, loss-to-follow up was even.

## *Comparisons*

### SJW versus Placebo

Seven studies, with the study by the Laakmann group split into two samples, were included in the comparison of SJW versus placebo in the treatment of mild to moderate depression.<sup>33</sup> The Cochran calculation revealed that there was not significant heterogeneity between included studies and that it would be appropriate to combine these studies into one meta-analysis. The pooled risk ratio for this comparison, calculated using a random effects model, was 1.3633 [95% CI: 1.0846 to 1.7138] with a p-value of 0.0079. These results demonstrate a statistically significant difference between the SJW and placebo groups, with an increased response rate in the SJW group. Please see figure 2 for a forest plot of the results from this meta-analysis.

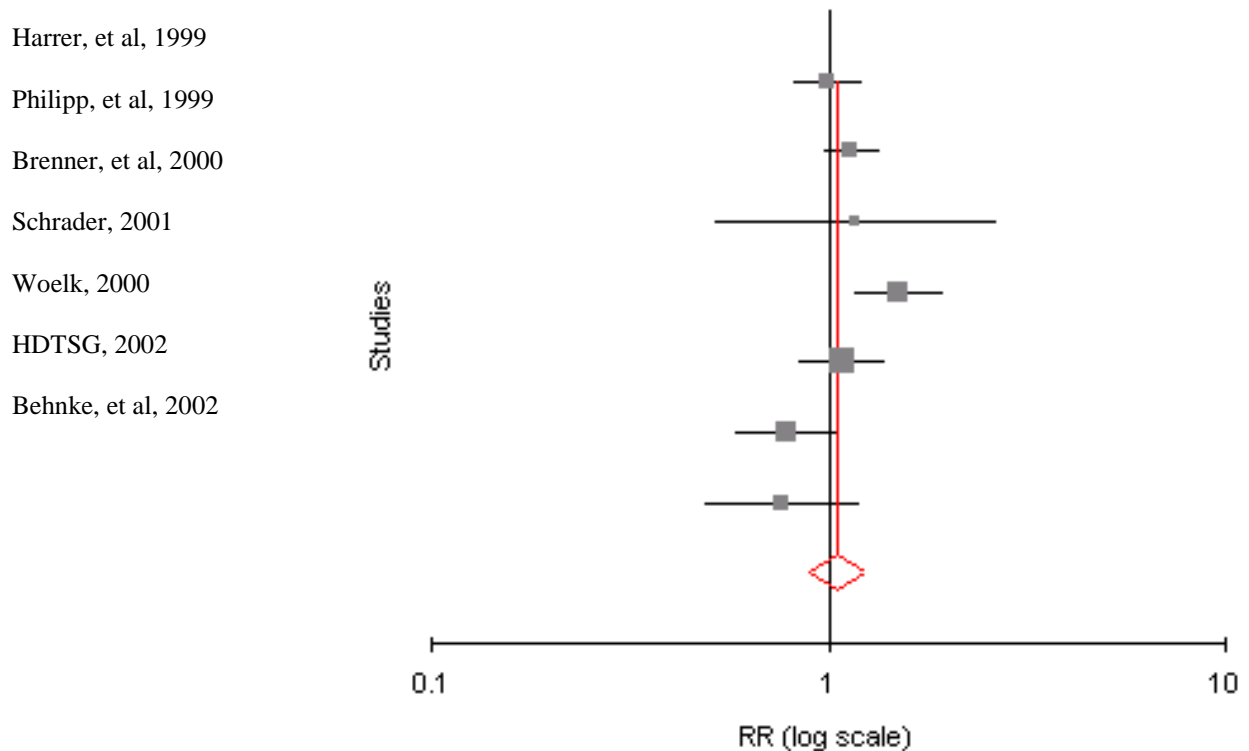


**Figure 2.** Forest plot produced through meta-analysis of included studies comparing SJW to placebo. The pooled risk ratio for this comparison, calculated using a random effects model, was 1.3633 [95% CI: 1.0846 to 1.7138] with a p-value of 0.0079. This comparison indicates that there is a significant difference in response to treatment with SJW versus placebo in favor of response. This demonstrates that SJW is more effective than placebo in treating mild to moderate depression.

### SJW versus Other Antidepressant

Seven studies were included in the comparison of SJW versus other antidepressant in the treatment of mild to moderate depression. The Cochran calculation revealed that there was not significant heterogeneity among included studies and that it would be appropriate to combine these studies into one meta-analysis. The pooled risk ratio for this comparison, calculated using a random effects model, was 1.0513 [95% CI: 0.8948 to 1.2352] with a p-value of 0.5428. These results do not demonstrate a statistically significant difference between the SJW and other antidepressant groups. In the studies analyzed, response rates of patients treated with SJW did

not differ significantly from response rates of those treated with other antidepressants. Please see figure 3 for the forest plot of the results from this meta-analysis.

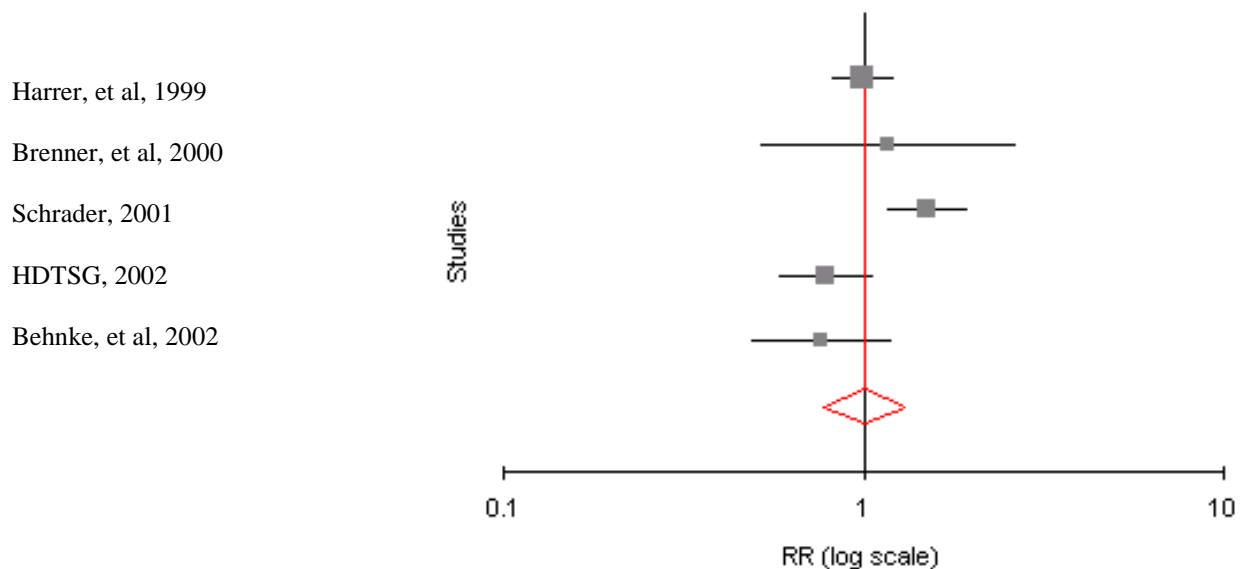


**Figure 3.** Forest plot produced through meta-analysis of included studies comparing SJW to other antidepressants. The pooled risk ratio for this comparison, calculated using a random effects model, was 1.0513 [95% CI: 0.8948 to 1.2352] with a p-value of 0.5428. This comparison indicates that there is no significant difference in response to treatment with SJW versus treatment with other antidepressants. This demonstrates that SJW is not different from, or is equivalent to, the other antidepressants studied in treating mild to moderate depression.

### SJW versus SSRI

A subset of five studies was separately analyzed to compare SJW versus SSRI in the treatment of mild to moderate depression. Since tricyclic antidepressants are not first-line treatment for depression, the studies that compared SJW to Imipramine were excluded from this

comparison.<sup>20, 40</sup> The Cochran calculation revealed that there was not significant heterogeneity among included studies and that it would be appropriate to combine these studies into one meta-analysis. The pooled risk ratio for this comparison, calculated using a random effects model, was 1.0066 [95% CI: 0.7694 to 1.317] with a p-value of 0.9615. These results do not demonstrate a statistically significant difference between the SJW and SSRI groups. In the studies analyzed, response rates of patients treated with SJW did not differ significantly from response rates of those treated with SSRI's. Please see figure 4 for a forest plot of the results from this meta-analysis.



**Figure 4.** Forest plot produced through meta-analysis of included studies comparing SJW to SSRI. The pooled risk ratio for this comparison, calculated using a random effects model, was 1.0066 [95% CI: 0.7694 to 1.317] with a p-value of 0.9615. This comparison indicates that there is no significant difference in response to treatment with SJW versus treatment with an SSRI. This demonstrates that SJW is not different from, or is equivalent to, the SSRI's studied in treating mild to moderate depression.

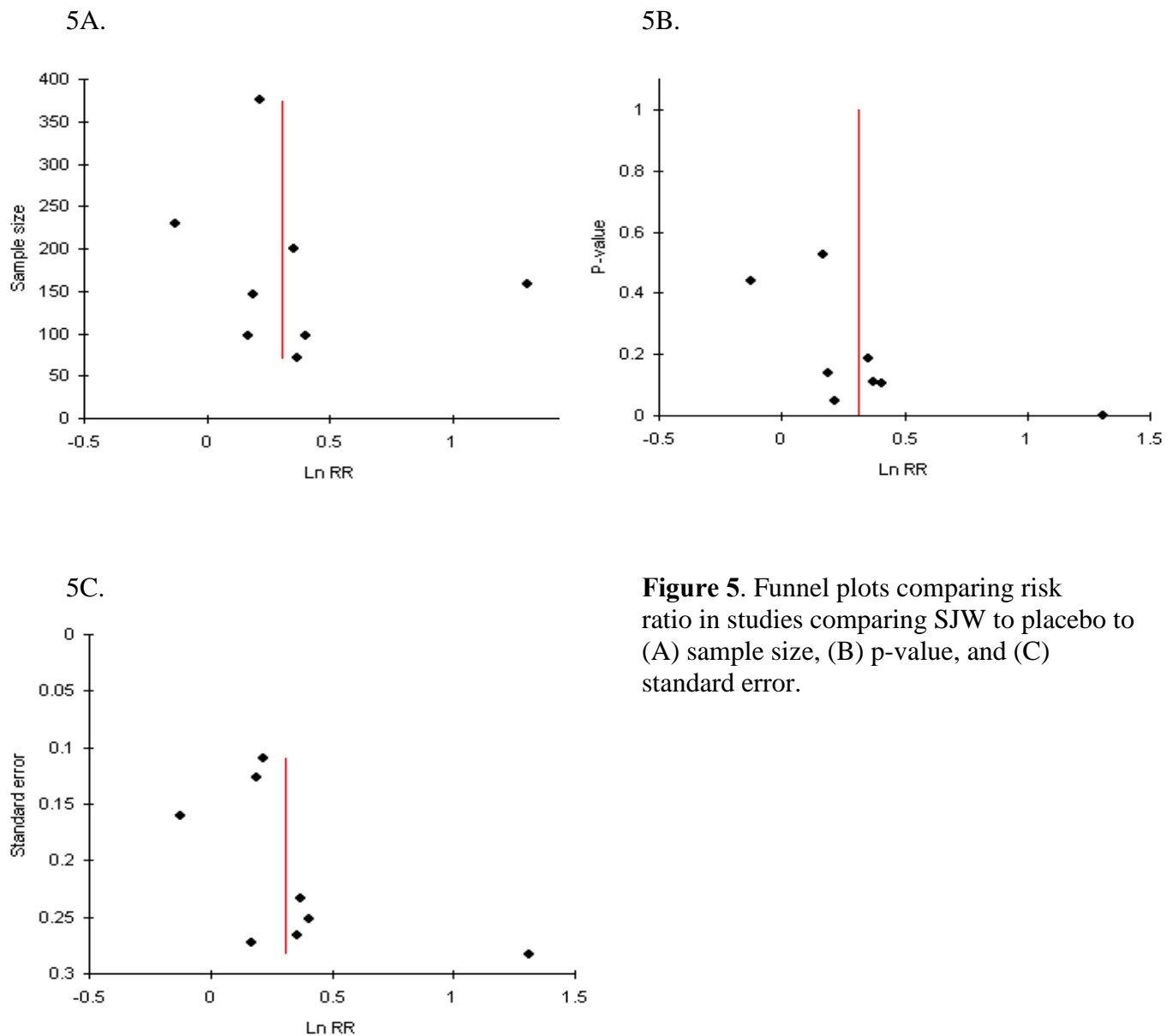
### Funnel Plots

Statistical analysis was used to identify sources of bias that may have affected the results of the set of studies on SJW included in this analysis. The Cochran calculation for homogeneity demonstrated the lack of significant heterogeneity between the studies included in the comparison of SJW versus placebo. The same is true for the comparisons of SJW versus other antidepressants and SJW versus SSRI. This initial calculation revealed that it was acceptable to combine the studies in each comparison for the purposes of producing a pooled outcome measure. Further analysis compared the treatment effects from each included study to certain other characteristics of these studies. Analyses produced three funnel plots for each comparison. Plots compared the natural log of each studies risk ratio to the sample size, p-value, and standard error. The solid line seen on each plot represents the pooled risk ratio for each comparison.

#### *SJW versus Placebo*

Funnel plots for this comparison revealed two outliers when individual studies risk ratios were compared to either sample size or level of significance (p-value). Schrader appeared to over-estimate the effects of SJW,<sup>9</sup> while the study performed by the HDTSG seemed to underestimate SJW effectiveness.<sup>21</sup> This result is demonstrated graphically by funnel plots 5A and 5B. In these plots, the study by HDTSG is further to the left of the remainder of included studies that are clustered around the pooled risk ratio. Points plotted to the left of the pooled risk ratio represent studies that favor the null hypothesis that SJW is not better than placebo. Schrader's study is represented by the point that is well to the right of the pooled risk ratio, in favor of a positive effect. A third funnel plot graphs individual study results versus standard

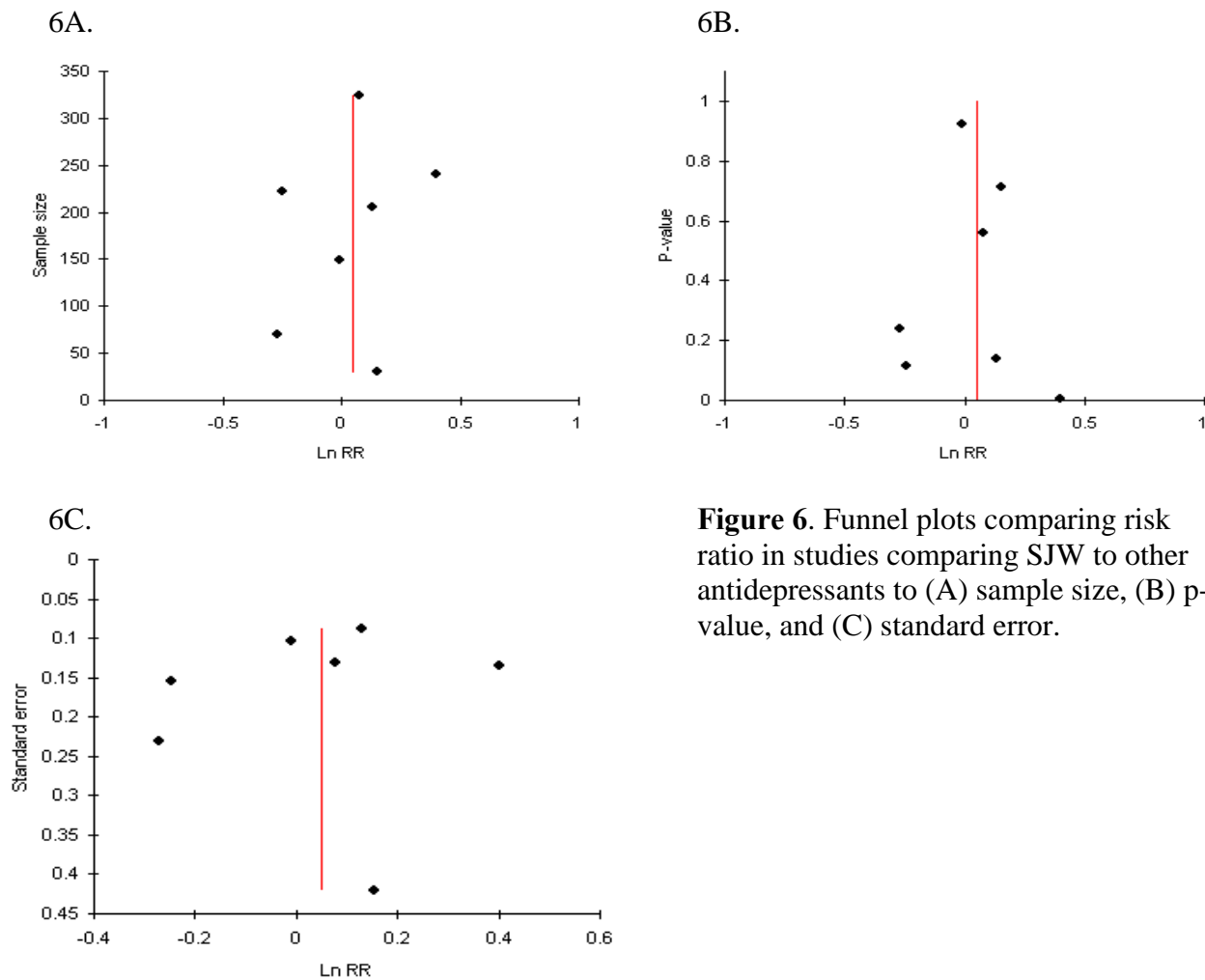
error. This comparison is designed to emphasize biases inherent in studies of smaller sample size. In this plot, seen in 5C, the studies by HDTSG and Schrader are outliers, with HDTSG again favoring lack of significant difference between SJW and placebo and Schrader favoring a positive effect for SJW. Please see figure 5 for funnel plots from this analysis.





### *SJW versus Other Antidepressants*

Funnel plots for this comparison demonstrate clustering of included studies around the solid line representing the pooled risk ratio regardless of sample size or p-value. Figure 6A shows that the sample size of an included study does not predict effect size. Figure 6B demonstrates that studies that support the null hypothesis that there is no difference between SJW and the other antidepressants studied and studies that reject this null hypothesis are equally distributed around the pooled risk ratio. Figure 6C shows that included studies are evenly distributed around the pooled risk ratio regardless of variance of the effect estimate within each included study. Please see figure 6 for funnel plots from this analysis.



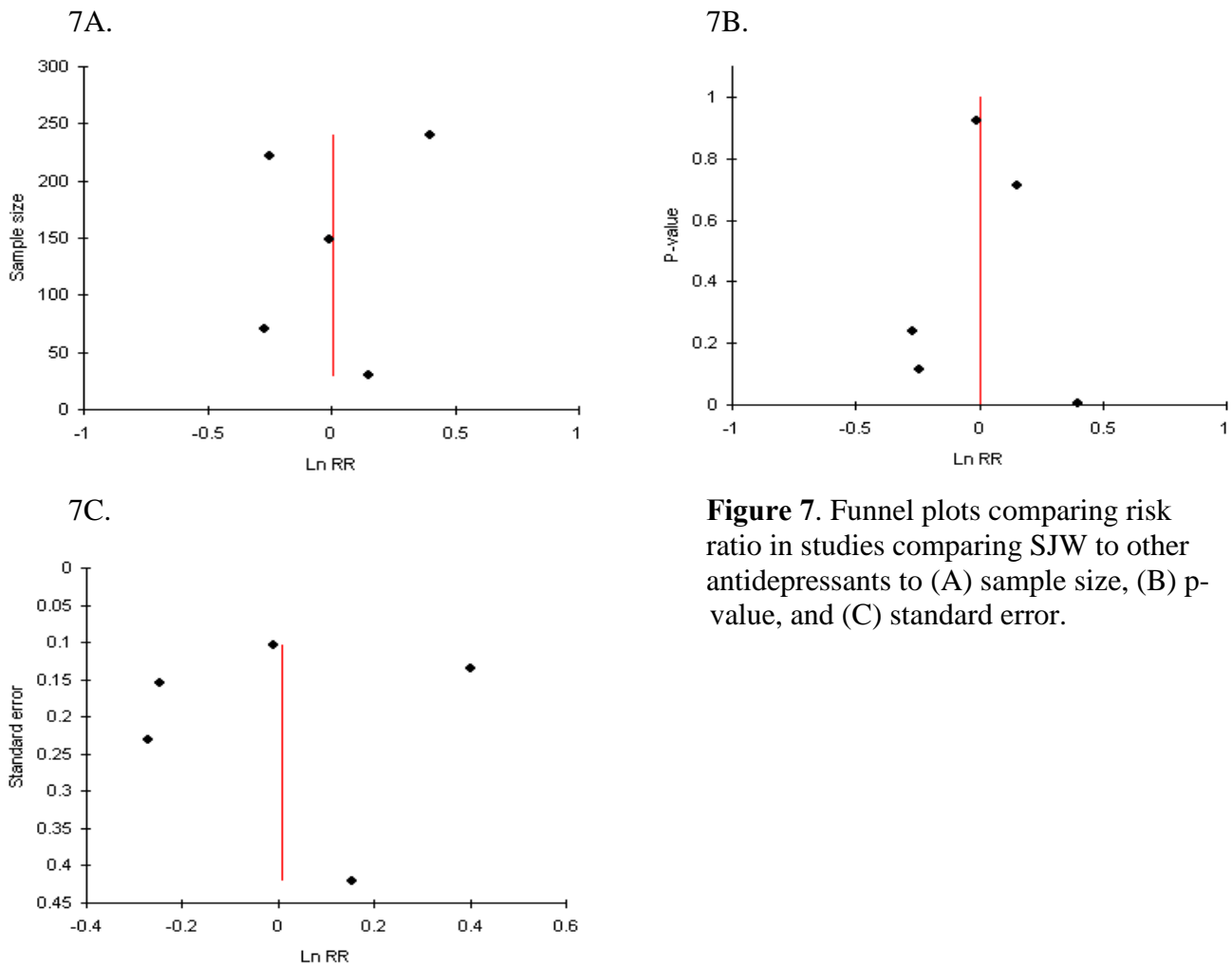
**Figure 6.** Funnel plots comparing risk ratio in studies comparing SJW to other antidepressants to (A) sample size, (B) p-value, and (C) standard error.

### *SJW versus SSRI*

Funnel plots for this comparison demonstrate clustering of included studies around the solid line representing the pooled risk ratio regardless of sample size or p-value. Figure 7A shows that the sample size of an included study does not predict effect size. Figure 7B demonstrates that studies that support the null hypothesis that there is no difference between SJW and the SSRI's studied and studies that reject this null hypothesis are equally distributed around the pooled risk ratio. Figure 7C shows that included studies are evenly distributed around

the pooled risk ratio regardless of variance of the effect estimate within each included study.

Please see figure 7 for funnel plots from this analysis.



**Figure 7.** Funnel plots comparing risk ratio in studies comparing SJW to other antidepressants to (A) sample size, (B) p-value, and (C) standard error.

## DISCUSSION

Before studying a novel antidepressant, one must review where pharmacotherapy fits in the treatment of depression. Initiating treatment for depression follows careful diagnosis and assessment of the patient for suicidal or self-injurious thoughts or plans. Use of one of several available diagnostic tools can assist clinical acumen in determining severity of depression. If a patient is severely depressed, then monotherapy with any antidepressant is inappropriate. If mildly to moderately depressed, however, it may be appropriate to begin therapy with an antidepressant. Research has demonstrated that monotherapy for depression is not as effective as combining pharmacotherapy and psychotherapy. Therefore, if practicing according to current best evidence, a physician would prescribe an antidepressant and refer the depressed patient to counseling once he or she had made the diagnosis of mild to moderate depression. In Germany, where several standardized SJW extracts are available and SJW has been approved by the appropriate government agency, SJW plays an important role as a cheap and safe part of outpatient treatment for depression.<sup>41</sup> In the US, however, commercially available SJW extracts are not standardized and are not currently regulated by the FDA. Under current conditions, SJW should not play a role in treatment for depression.

Like many herbal medications, however, SJW has found popularity in the US marketplace. Several sources report that sales of SJW products have increased over the past decade.<sup>41-43</sup> As with other herbal supplements, many patients do not report their use of supplements to their physicians unless specifically asked. As a result, it is difficult to assess the potential for important drug interactions. Since herbal supplements do not go through the same FDA approval process as synthetic pharmaceuticals, it is difficult to predict what adverse events

are possible with use of SJW. Over recent years, both adverse events and important drug interactions have been coming to light through post-market reports on SJW products in the US. Adverse events include serotonin syndrome when SJW is taken at the same time as a monoamine oxidase inhibitor or a SSRI<sup>44</sup> as well as several effects resulting from activation of the cytochrome p450 enzyme complex. Such p450 effects include lowering of serum cyclosporine concentrations leading to graft rejection in transplant patients,<sup>45</sup> lowering of hormone levels leading to breakthrough bleeding in women on oral contraceptives to control menstrual bleeding,<sup>46</sup> and decreased INR in patients taking coumadin<sup>47</sup> or lowering serum digoxin concentration in cardiac patients.<sup>48</sup> Each of these examples represents an easily preventable adverse event. Such events are only preventable, however, if the provider is aware of the potential of SJW for such interactions and that the patient is taking SJW. If the FDA treated herbal supplements, including SJW, as any other pharmaceutical, it would be possible to more closely observe and regulate their use. More specifically, pre-market research would uncover significant potential adverse events and alert physicians so that they may properly counsel patients considering taking SJW.

This analysis set out to answer whether recent research has shown SJW to be effective in treating mild to moderate depression. Statistical analysis did not identify any significant heterogeneity in the studies included in the three comparisons. While it may have been appropriate to combine the studies used for each comparison, the resulting pooled outcomes measures must still be critically reviewed. To begin, the pooled outcomes measures for this analysis indicate that SJW is significantly more effective than either placebo and equivalent to other antidepressants in treating mild to moderate depression. Closer analysis reveals that there

may be more questions raised by this analysis than answered through the meta-analysis of included studies.

Studies of SJW have been performed in two countries with decidedly differing views on its utility as an antidepressant. European nations have a more accepting view towards herbal remedies, with SJW being prescribed more often than Prozac in Germany.<sup>14</sup> The medical community in the United States, however, is less accepting of herbal medicines. This attitude is reflected in the conclusions of both studies comparing SJW to placebo or other antidepressants conducted in the US and published in JAMA. Shelton and associates found that SJW was not more effective than placebo in treatment of mild to moderate depression.<sup>28</sup> The HDTSG also found that SJW was not more effective than placebo and added that SJW was not equivalent to Sertraline in treating mild to moderate depression.<sup>21</sup> These investigators evaluated past studies of SJW and determined that past studies lacked the methodological rigor to give credence to past conclusions that SJW was effective in treating depression.

Shelton, et al and HDTSG took extra steps to ensure meaningful conclusions. These measures included using DSM-IV criteria to define mild to moderate depression at baseline, using intention to treat analysis, sufficient study length to detect differences between treatments, and using an appropriate dose of any other antidepressant being compared to SJW. Applying DSM-IV criteria to subjects at baseline allows for more accurate clinical diagnosis and inclusion of only mild to moderately depressed subjects. Other studies have used ICD-10 criteria, which may mask some improperly diagnosed subjects, including those whose depressed mood may lift spontaneously regardless of treatment. Both of these studies were 8 weeks in length. A strong argument could be made that study length should be longer to allow for real differences to

manifest between study groups. Eight weeks is generally the shortest period considered to be a legitimate trial for an antidepressant. The dose of Sertraline used in the HDTSG study, 50 to 100-mg, is the average therapeutic dose for this SSRI.<sup>49</sup> Both of these studies used sufficient sample size to detect differences between groups.

In comparison to the studies by Shelton, et al and HDTSG, other studies in this analysis have not met the proposed standards for methodological rigor. The most common weaknesses of the other studies includes failing to use DSM-IV criteria to define mild to moderate depression, duration too short to detect differences, differential loss to follow-up, and insufficient sample size. These other studies also share some other characteristics. Except for studies by Shelton and associates,<sup>28</sup> HDTSG,<sup>21</sup> and Brenner, et al,<sup>38</sup> studies included in this analysis were mostly performed in Germany and published in lesser-known journals. Even Brenner's study was published in a lesser-known journal, whereas the studies by Shelton's group and the HDTSG were published in JAMA, widely considered one of the most respected journals in medicine. Due to an assumed more rigorous peer review process, these two studies may reasonably be considered to carry more subjective and objective weight when evaluating the literature on this subject published since 1998.

Only two studies in this analysis employed a design sufficient to control for the placebo effect. The placebo effect confounds studies of antidepressants by making it more likely to conclude a novel therapy is more effective than an SSRI (if studied without placebo) or not as effective as placebo (if studied alone versus placebo).<sup>50</sup> Only two studies included in this analysis have included both placebo and other antidepressant groups in comparison to SJW. Philipp, et al compared SJW to both placebo and Imipramine. This group concluded that SJW

was more effective than placebo and as effective as Imipramine.<sup>20</sup> This analysis concluded that this was a fair study though its conclusions are weakened by its use of a dose of Imipramine well below the average therapeutic dose.<sup>49</sup> Its conclusions are made even less useful since Imipramine is no longer first line pharmacotherapy for depression.

HDTSG more successfully employs adequate design to reduce the placebo effect. In fact, this study demonstrates the model study design and methodology that will allow future studies to accurately assess the effectiveness of SJW. This model design includes a placebo group to test the effectiveness of SJW versus no treatment. Using such methods as a standard pill size across all groups should adequately blind this placebo group. The study should be from 8 to 12 weeks to give an adequate trial of medication and time to adjust dosages if necessary. Since SSRI is currently the first line pharmacotherapy for depression, SJW should be compared to an adequate dose of a SSRI. Standard formulations of both SJW and SSRI should be used. These formulations of SJW should reflect what is currently available in the marketplace. Future studies should also meet the conditions for methodological rigor set forth by the HDTSG.<sup>21</sup> According to this analysis, there is one study that adequately compares SJW to placebo and one study that adequately compares SJW to an SSRI. More work is needed before any final conclusions can be drawn about the effectiveness of SJW in treating mild to moderate depression.

This review was limited by its restriction to English language journals, to published data and to studies published from 1998 to present. There may be unpublished data either proving or disproving the effectiveness of SJW. Inclusion of foreign language journals may reveal more studies appropriate for this review, especially since a majority of studies found were performed in Germany. Studies prior to 1998 were excluded secondary to methodological weakness.



Reviewing these studies through a methodologically rigorous eye may identify some studies that would help shed light on SJW effectiveness when compared to placebo. The small number of studies included weakened the meta-analysis. Repeating this review at a future date after more studies meeting criteria for methodological rigor have been performed would increase the strength of the meta-analysis.

## REFERENCES CITED

1. Murray CJL, Lopez AD. The Global Burden of Disease: A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020, Summary. <http://www.who.int/msa/mnh/ems/dalys/intro.htm>. Accessed July 2, 2004.
2. National Institutes of Mental Health. The numbers count: Mental disorders in America. <http://www.nimh.nih.gov/>. Accessed July 2, 2004.
3. Hirschfeld RMA, Keller MB, Pancino S, et al. The National Depressive and Manic-Depressive Association consensus statement on the undertreatment of depression. *JAMA*. 1997; 277:333-340.
4. Ernst E, Rand JJ, Barnes J, Stevinson C. Complementary therapies for depression: An overview. *Arch Gen Psychiatry*. 1998; 55:1026-1032.
5. Astin JA. Why patients use alternative medicine: results of a national study. *JAMA*. 1998; 279:1548-1553.
6. Vincent C, Furnham A. Why do patients turn to complementary medicine? An empirical study. *Br J Clin Psychol*. 1996; 35:37-48.
7. Gupta RK, Moller H-J Moller. St. John's Wort: An option for the primary care treatment of depressive patients? *Eur Arch Psychiatry Clin Neurosci*. 2003; 253:140-148.
8. Gaster B, Holroyd J. St. John's wort for depression: A systematic review. *Arch Intern Medicine*. 2000; 160:152-156.
9. Schrader E, Meier B, Brattstrom A. Hypericum treatment of mild-moderate depression in a placebo-controlled study: A prospective, double-blind, randomized, placebo-controlled, multicentre study. *Human Psychopharmacol*. 1998; 13:163-169.
10. Cirigliano, MD; for UpToDate. Clinical use of St. John's wort. <http://www.utdol.com>. Accessed April 28, 2004.
11. Kim HL, Streltzer J, Goebert D. St John's wort for depression: A meta-analysis of well-defined clinical trials. *J Nervous Mental Dis*. 1999; 187(9):532-538.
12. Delito J, Beyer D. The scientific, quasi-scientific and popular literature on the use of St. John's wort in the treatment of depression. *J Affective Dis*. 1998; 51:345-351.
13. Whiskey E, Werneke U, Taylor D. A systematic review and meta-analysis of Hypericum perforatum in depression: A comprehensive clinical review. *Int Clin Psychopharm*. 2001; 16:239-252.

14. Linde K, Ramirez G, Mulrow CD, Pauls A, Widenhammar W, Melchart D. St. John's wort for depression: An overview and meta-analysis of randomized clinical trials. *BMJ*. 1996; 313:353-358.
15. Bessell TL, Anderson JN, Silagy CA, Sansom LN, Hiller JE. Surfing, self-medicating and safety: Buying non-prescription and complementary medicines via the Internet. *Qual Saf Health Care*. 2003; 12:88-92.
16. Morris CA, Avorn J. Internet marketing of herbal products. *JAMA*. 2003; 290:1505-1509.
17. Walsh BT, Seidman SN, Sysko R, Gould M. Placebo response in studies of major depression: Variable, substantial, and growing. *JAMA*. 2002; 287:1840-1847.
18. Kupfer DJ, Frank E. Placebo in clinical trials for depression: Complexity and necessity. *JAMA*. 2002; 287:1853-1854.
19. Charney DS, Nemeroff CB, Lewis L, et al. National Depressive and Manic-Depressive Association Consensus Statement on the use of placebo in clinical trials of mood disorders. *Arch Gen Psychiatry*. 2002; 59:262-270.
20. Philipp M, Kohnene R, Hiller K-O. Hypericum extract versus imipramine or placebo in patients with moderate depression: Randomized multicentre study of treatment for eight weeks. *BMJ*. 1999; 319:1534-1538.
21. Hypericum Depression Trial Study Group. Effect of *Hypericum perforatum* (St. John's wort) in major depressive disorder: A randomized control trial. *JAMA*. 2002; 287:1807-1814.
22. Dietary Supplement Health and Education Act. Pub L No. 103-417, 108 Stat 4325. 1994.
23. US Food and Drug Administration. Regulations on statements made for dietary supplements concerning the effect of the product on the structure or function of the body. Final rule, 65 *Federal Register* 1000. 2000.
24. Federal Trade Commission. *Dietary supplements: An advertising guide for industry*. <http://www.ftc.gov/bcp/online/pubs/buspubs/dietsupp.htm>. Accessed April 15, 2004.
25. DeAngelis CD, Fontanarosa PB. Drugs alias dietary supplements. *JAMA*. 2003; 290:1519-1520.
26. US Food and Drug Administration. *Claims that can be made for conventional foods and dietary supplements*. <http://www.cfsan.fda.gov/~dms/hclaims.html>. Accessed April 15, 2004.

27. US Food and Drug Administration. *Overview of dietary supplements*. <http://www.cfsan.fda.gov/~dms/ds-oview.html>. Accessed April 15, 2004.
28. Shelton RC, Keller MB, Gelenberg A, et al. Effectiveness of St. John's wort in major depression: A randomized control trial. *JAMA*. 2001; 285:1978-1986.
29. American Psychiatric Association. *Depressive Disorders*. In: Diagnostic and Statistical Manual of Mental Disorders. 4<sup>th</sup> ed, Text Revision. Washington, DC: American Psychiatric Association. 2000.
30. Zimmerman M, Posternak MA, Chelminski I. Is the cutoff to define remission on the Hamilton rating scale for depression too high? *J Nervous Mental Disease*. 2005; 193(3):170-175.
31. Bax L, Yu LM, Ikeda N, Tsuruta H, Moons KGM. MIX: Comprehensive free software for meta-analysis of causal research data. Version 1.7. <http://mix-for-meta-analysis.info>, 2008.
32. Bax L, Yu LM, Ikeda N, Tsuruta H, Moons KGM: Development and validation of MIX: comprehensive free software for meta-analysis of causal research data. *BMC Medical Research Methodology* 2006, 6(50).
33. Laakmann, G, Dienel A, Kieser M. Clinical significance of hyperforin for the efficacy of hypericum extracts on depressive disorders of different severities. *Phytomedicine*. 1998; 5:435-442.
34. Kalb R, Trautmann-Sponsel RD, Kieser M. Efficacy and tolerability of Hypericum extract WS 5572 versus placebo in mildly depressed patients: A randomized double-blind multicenter clinical trial. *Pharmacopsychiatry*. 2001; 34:96-103.
35. Lecrubier Y, Clerc G, Didi R, Kieser M. Efficacy of St. John's wort extract WS 5570 in major depression: A double-blind, placebo-controlled trial. *Am J Psychiatry*. 2002; 159:1361-1366.
36. Harrer G, Schmidt U, Kuhn U, Biller A. Comparison of equivalence between the St. John's wort extract LoHyp-57 and fluoxetine. *Arzneimittelforschung*. 1999; 49:289-96.
37. Brenner R, Azbel V, Madhusoodanan S, Pawlowska M. Comparison of an extract of hypericum (LI 160) and sertraline in the treatment of depression: A double-blind, randomized pilot study. *Clinical Therapeutics*. 2000; 22:411-419.
38. Schrader E. Equivalence of St. John's wort (Ze 117) and fluoxetine: A randomized, controlled study in mild-moderate depression. *Int Clin Psychopharm*. 2001; 16:61-68.
39. Woelk H. Comparison of St. John's wort and imipramine for treating depression: Randomized controlled trial. *BMJ*. 2000; 321:536-539.

40. Behnke K, Jensen GS, Graubaum HJ, Gruenwald J. Hypericum perforatum versus fluoxetine in the treatment of mild to moderate depression. *Advances in Therapy*. 2002; 19:43-52.
41. Bilia AR, Gallori S, Vinicieri FF. St. John's wort and depression: Efficacy, safety and tolerability – an update. *Life Sciences*. 2002; 70(26):3077-3096.
42. Bessell TL, Anderson JN, Silagy CA, Sansom LN, Hiller JE. Surfing, self-medicating and safety: Buying non-prescription and complementary medicines via the Internet. *Qual Saf Health Care*. 2003; 12:88-92.
43. Morris CA, Avorn J Internet Marketing of Herbal Products. *JAMA*. 2003; 290:1505-1509.
44. Lantz MS, Buchalter E, Giambanco V. St. John's wort and antidepressant interactions in the elderly. *J Geriatric Psych and Neuro*. 1999; 12:7-10.
45. Barone GW, Gurley BJ, Ketel BL, Lightfoot ML, Abul-Ezz SR. Drug interaction between St. John's wort and cyclosporine. *Annals of Pharmacotherapy*. 2000; 34(9):1013-1016.
46. Bon S, Hartmann K, Kuhun M, Joanniskraut ein Enzyminduktor? *Schweitzer Apothekerzeitung*. 1999; 16:535-536.
47. Yue QY, Bergquist C, Gerden B. Seven cases of decreased effect of warfarin during concomitant treatment with St. John's wort. *Lancet*. 2000; 355:576-577.
48. Johne A, Brockmoller J, Bauer S, Maurer A, Lanheinrich M, Roots I. Pharmacokinetic interaction of digoxin with an herbal extract from St. John's wort (*Hypericum perforatum*). *Clinical Pharmacology and Therapeutics*. 1999; 66(4):338-345.
49. Aronson SC, Ayres VE. Depression: A treatment algorithm for the family physician. *Hospital Physician*. 2000: 21-44.
50. Kupfer DJ, Frank E. Placebo in Clinical Trials for Depression: Complexity and Necessity. *JAMA*. 2002; 287:1853-1854.